Is proton pump inhibitor testing an effective approach to diagnose gastroesophageal reflux disease in patients with noncardiac chest pain: a meta-analysis

CRD summary
This review addressed the usefulness of initial treatment with a proton-pump inhibitor (PPI) as a method of diagnosing gastroesophageal reflux disease (GERD) in patients with noncardiac chest pain. The review was well conducted and clearly reported. The authors’ conclusion, that initial treatment with PPI is a valid method of testing for GERD, is likely to be reliable.

Authors’ objectives
To evaluate the accuracy of a proton-pump inhibitor (PPI) treatment for the diagnosis of gastroesophageal reflux disease (GERD)-related noncardiac chest pain (NCCP).

Searching
MEDLINE, EMBASE, CINAHL, and the Cochrane Controlled Trials Register were searched from inception to May 2004. The search terms were reported and no language restrictions were applied. The proceedings of four international gastroenterological meetings were also searched. The reference lists of all relevant review articles and primary studies retrieved were checked for additional studies.

Study selection
Study designs of evaluations included in the review
Only randomised, placebo-controlled trials were eligible for inclusion.

Specific interventions included in the review
Studies comparing PPI with placebo were eligible for inclusion.

Reference standard test against which the new test was compared
Studies in which the diagnosis was confirmed by endoscopy and/or 24-hour oesophageal pH monitoring were eligible for inclusion.

Participants included in the review
Studies of adult patients with recurrent chest pain and with no known cardiac abnormalities were eligible for inclusion. The mean age of the participants was 54.4 years (standard deviation 6.1) and 60.4% were male (derived from 5 of the 6 included studies).

Outcomes assessed in the review
The included studies were required to report sufficient information to derive 2x2 contingency tables. The calculated outcome measures reported in the review were the sensitivity, specificity, positive predictive value, negative predictive value and diagnostic odds ratio (DOR).

How were decisions on the relevance of primary studies made?
Three reviewers independently conducted the literature search. Titles and abstracts were screened for relevance before retrieval of the full articles. Where there was uncertainty, the full article was retrieved.

Assessment of study quality
Criteria published on behalf of the Cochrane Screening and Diagnostic Tests Methods Group were used to assess the quality of the included studies. These criteria considered: the spectrum of disease; the reference standard used to confirm diagnosis; definition of disease; blinding of measurements; execution of the index test; explicit definition of
the improvement of symptoms and reporting of the index test cut-off; and reporting of demographic information and sampling strategy (see Other Publications of Related Interest no.1). The authors did not specifically describe how quality was assessed, or how many reviewers performed the quality assessment.

Data extraction
Three reviewers extracted the data independently using a predefined spreadsheet. Any disagreements were resolved by discussion to reach consensus.

Methods of synthesis
How were the studies combined?
Pooled estimates of the sensitivity, specificity and DOR, along with their 95% confidence intervals (CIs), were calculated using a random-effects model. Summary receiver operating characteristic (sROC) curves were estimated, using the method of Littenberg and Moses (see Other Publications of Related Interest no.2), to take account of the effects of varying diagnostic threshold upon test accuracy. sROC curves were produced for groups of studies with comparable reference standards, methods of defining symptom improvement, and natural log DOR (i.e. where it was reasonable to assume a degree of clinical and statistical homogeneity). Areas under sROC curves (AUC) were also calculated; sROC curves and AUC were not extrapolated beyond existing data points. Analyses were conducted, both unweighted and weighted by inverse variance. The difference between the DOR of PPI and placebo was assessed using a Wilcoxon paired-sample test.

How were differences between studies investigated?
Between-study heterogeneity was assessed statistically using a chi-squared test and visually using plots of sensitivity, specificity and DOR with their 95% CI. Where visual heterogeneity was observed, possible sources of clinical heterogeneity were sought. The possible influence of covariates (e.g. elements of study quality, prevalence of NCCP, treatment duration) on accuracy was assessed using regression analyses, by extending the sROC model.

Results of the review
Six trials (n=220) were included in the review.

Study quality.
Five of the six included studies were double blind and one was single blind; five were crossover studies. The diagnosis of NCCP was clearly described in five studies, while the nature of the NCCP in study patients was clearly described in four.

Comparative diagnostic accuracy.
No between-study statistical heterogeneity was identified for PPI (P=0.95) or placebo (P=0.17).

The pooled estimates for the sensitivity and specificity of a PPI diagnostic test were 0.80 (95% CI: 0.71, 0.87) and 0.74 (95% CI: 0.64, 0.83), respectively. The comparative estimates for the placebo group were 0.19 (95% CI: 0.12, 0.29) and 0.77 (95% CI: 0.62, 0.87), respectively.

The PPI test had significantly higher overall diagnostic accuracy for diagnosing GERD in patients with NCCP than placebo: the pooled DOR was 19.35 (95% CI: 8.54, 43.84), compared with 0.61 (95% CI: 0.20, 1.86, P=0.03).

Regression analyses indicated that type and year of publication, study design, prevalence of NCCP, reference standard used and treatment duration had no significant effect on estimates of test accuracy.

Authors’ conclusions
PPI treatment, as a diagnostic test, has adequate sensitivity and specificity to detect GERD in patients with NCCP, and could be used as an initial approach by general practitioners.
CRD commentary
This was a well-conducted review that addressed a clearly stated research question and reported appropriate inclusion criteria. A thorough literature search was conducted, no language restrictions were applied, and efforts were made to identify unpublished studies. Methods appropriate to the assessment of diagnostic studies were used to examine the methodological quality of the included studies, and appropriate measures were taken to avoid the introduction of error and bias during the review process. Statistical analyses were rigorous, appropriately applied, and clearly reported. The authors' conclusions are supported by the evidence presented.

Implications of the review for practice and research
Practice: The authors suggested that treatment with high dosage PPI for up to 4 weeks has acceptable sensitivity and specificity. They further stated that it could be used by general practitioners in patients with NCCP and suspected GERD; if more than a 50% reduction in symptom scores is achieved, then the chance of GERD is greatly increased and PPI treatment should be continued.

Research: The authors stated that further well-designed, adequately powered studies are required to support the findings of this analysis.

Funding
Gastroenterological Research Fund, University of Hong Kong.

Bibliographic details

PubMedID
15956000

DOI
10.1001/archinte.165.11.1222

Original Paper URL
http://archinte.ama-assn.org

Other publications of related interest

Indexing Status
Subject indexing assigned by NLM

MeSH
Anti-Ulcer Agents /therapeutic use; Chest Pain /etiology; Diagnostic Techniques, Digestive System; Gastroesophageal Reflux /complications /diagnosis /drug therapy; Humans; Proton Pump Inhibitors; Sensitivity and Specificity

AccessionNumber
12005008358

Date bibliographic record published

Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.